



Compositional modelling of signalling pathways in timed concurrent constraint programming

Davide Chiarugi, Moreno Falaschi, Carlos Olarte, Catuscia Palamidessi

► To cite this version:

Davide Chiarugi, Moreno Falaschi, Carlos Olarte, Catuscia Palamidessi. Compositional modelling of signalling pathways in timed concurrent constraint programming. First ACM International Conference on Bioinformatics and Computational Biology (BCB '10)., Aug 2010, Niagara Falls, New York, United States. pp.414-417, 10.1145/1854776.1854843 . hal-00548213

HAL Id: hal-00548213

<https://hal.science/hal-00548213>

Submitted on 20 Dec 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Compositional Modelling of Signalling Pathways in Timed Concurrent Constraint Programming*

Davide Chiarugi
Dip. Scienze Matematiche e
Informatiche
Università di Siena, Italy
chiarugi3@unisi.it

Moreno Falaschi
Dip. Scienze Matematiche e
Informatiche Università di
Siena, Italy
moreno.falaschi@unisi.it

Carlos Olarte
Dept. Computer Science
Universidad Javeriana.
Cali, Colombia
carlosolarte@puj.edu.co

Catuscia Palamidessi
INRIA and LIX, Ecole
Polytechnique, France
catuscia@lix.polytechnique.fr

ABSTRACT

The biological data regarding the signalling pathways often consider single pathways or a small number of them. We propose a methodology for composing this kind of data in a coherent framework, in order to be able to investigate a bigger number of signalling pathways. We specify a biological system by means of a set of stoichiometric-like equations resembling the essential features of molecular interactions. We represent these equations by a timed concurrent constraint (**ntcc**) language, which can deal with partial information and the time for a reaction to occur. We describe a freely available prototypical implementation of our framework.

1. INTRODUCTION

Cellular signalling pathways consist in groups of interacting proteins belonging to various functional classes, e.g. receptors, adaptor proteins and kinases [4]. Each pathway is specialized in sensing and transducing particular environmental signals such as growth factors, hormones, cytokines, light and nutrients. Misfunctioning or defects in signalling pathways are often associated with important diseases e.g. diabetes, immune disorders and cancer [4, 7]. In living cells individual signalling pathways do not act in isolation. Rather, they cross talk due to the presence of components shared by different pathways or positive and negative feedback loops that form complex interaction networks. Thus, the overall behavior of a cell embedded in a given environment should be seen as the integrated response to a variety of sensed signals coming from the outside [4, 1]. Understanding the dynamics underlying the functioning of whole

cellular signalling networks is a task that is far from being accomplished. Unfortunately, nowadays wet-lab techniques are still not adequate to cope with this kind of “systemic” investigation. Moreover, the available information results to be heterogenous as it is obtained through different experimental techniques and comes from variable contexts. So, it results particularly difficult to organize the existing information in a coherent framework taking also into account the dynamics of the interactions, in order to gain knowledge at the network level. To address the problem of data integration and of heterogeneity, we present a new methodology for modelling molecular interaction networks, which allows to organize the existing knowledge in a compositional fashion. We aim at exploiting this method for describing and simulating the behavior of a big number of signalling pathways. Our modelling technique takes into account the fact that several features of the biological system may be undetermined or affected by measuring mistakes [5]. Our language allows us to represent partial information and to model explicitly the time which is necessary for a reaction to occur. In this paper we use our methodology by modeling and simulating some well known signalling pathways, namely the TNF α , EGF, and insulin signalling pathways.

2. CONCURRENT CONSTRAINT CALCULI

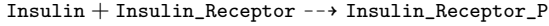
Concurrent system, i.e. systems whose components act simultaneously and potentially interact with each other, are ubiquitous in several domains and applications, as in biological and physical systems, mobile systems, Internet, etc. Due to their complex forms of interaction, concurrent systems are difficult to specify and reason out. In computer science, process calculi have been proposed to rise this challenge. They provide a mathematical language where concurrent systems can be modelled and their behaviour can be formally described and predicted. Timed Concurrent Constraint Programming (**ntcc**) [13] is a simple but powerful model for concurrency aiming at modeling reactive systems, i.e., those whose components interact among each other and with the environment. Here, interactions are modeled by means of constraints accumulated in a common store. A constraint c , e.g., $x \geq 10$, represents a piece of (partial) information upon which processes may act. Processes in **ntcc** can change the state of the system by telling information, and

*This work has been partially supported by the Italian MUR under grant RBIN04M8S8, FIRB project, Internationalization 2004.

synchronize by asking information: A process $\text{tell}(c)$ adds c to the store, thus making it available to the other processes. A process $\sum_{i \in I} \text{when } c_i \text{ do } P_i$ non-deterministically chooses one of the P_i whose corresponding guard c_i can be deduced from the store. The chosen alternative, if any, precludes the others. A process $P \parallel Q$ denotes P and Q running concurrently possibly “communicating” via the common store. Processes in **ntcc** can be also executed along time-units or time-intervals: A process $\text{next } P$ delays the execution of P to the next time-unit, and the *replication* $!P$ means unboundedly many copies of P but one at a time. We refer the reader to [13] for a detailed account of the syntax and the operational semantics of the **ntcc** calculus.

3. A COMPOSITIONAL MODEL

A biochemical pathway can be represented as a set of interconnected biochemical reaction in which the product of a reaction is the reactant of another one. A signaling pathway is a biochemical pathway supporting the transduction of signals in a biological cell: molecules present in the external environment can affect the cellular behaviour linked to a specific receptor on the cell membrane and triggering the biochemical reactions of the receptor-dependent signaling pathways. For example, the equation



represents the reaction of one insuline molecule and its receptor (IR). This reaction causes the (auto)phosphorilation of the receptor molecule, and triggers the following events of the signal transduction pathway. Here we abstracted away various details regarding chemical features, such as reaction rate dependence from the temperature. These approximations are consistent with the *in-vitro* experiments, where the temperature is kept constant. The reaction rate will be considered at simulation time as stochastic parameters. In general, we shall represent the signaling pathway by a finite set of equations r_1, \dots, r_m of the form

$$r_j : \sum_i a_i^j A_i \dashrightarrow \sum_i b_i^j B_i \quad (1)$$

The constants a_1^j, \dots, a_n^j and b_1^j, \dots, b_n^j are the stoichiometric coefficients. Therefore, $a_1^j A_1, a_2^j A_2, \dots, a_n^j A_n$ are reactants that interact (and are consumed) yielding to the products $b_1^j A_1, b_2^j A_2, \dots, b_n^j A_n$. We also assume that each equation r_j has associated a duration $\text{dur}(r_j)$ defining the number of time-units required for that reaction to produce the components on the right-hand side.

Our idea is to represent this set of equations as **ntcc** processes. This will provide a runnable specification of the system in which we can formally specify the time required for a reaction to occur. Furthermore, due to the constraint-based nature of **ntcc**, and the ability of dealing with partial information, we will be able to specify systems where we do not have a complete description of it (e.g., if we only have a partial information about the concentration of the reactants involved). Moreover, models specified in this language can be easily composed due to the parallel composition operator, thus yielding the description of more complex phenomena. Finally, another advantage of using **ntcc**, is that it is endowed with a logic semantics [13], thus providing a valuable tool for the verification of properties of the modelled system.

3.1 The model in **ntcc**

The model we propose consists of three different components: A process to *choose* the rule (equation) to be applied in each time-unit; the processes modelling the *equations* and the processes changing the *state* of the concentrations of each reactant/product according to the equation applied.

We assume a set of equations as in Equation 1. In the sequel, the variables x_1, \dots, x_n represent the current concentrations of the components A_1, \dots, A_n respectively.

The process **choose** below, selects non-deterministically one of the equations ($eq = j$) s.t. the current concentration of each component is higher than the reactant necessities for the equation to take place (e.g., $x_i^j \geq a_i^j$).

$$\text{choose} \stackrel{\text{def}}{=} \sum_{j=1}^m \text{when } x_1^j \geq a_1^j \wedge \dots \wedge x_n^j \geq a_n^j \text{ do tell}(eq = j)$$

Improvements here can be done by adding the stochastic information about the reaction rate of each equation (if it is known). In this case, the non-deterministic operator is replaced by a probabilistic choice (see e.g., [11]).

A reaction (equation) r_j is modelled by a process binding the variables x_{i-}^j and x_{i+}^j to a_i^j and b_i^j respectively. This variables determine how the concentration of the component A_i must be affected due to the application of the equation: a_i^j units are consumed and b_i^j units are produced.

$$\begin{aligned} \text{equation}_j \stackrel{\text{def}}{=} & \text{when } eq = j \text{ do} \\ & \text{next tell}(x_{1-}^j = a_1^j \wedge \dots \wedge x_{n-}^j = a_n^j) \parallel \\ & \text{next } \text{dur}(r_j) \text{ tell}(x_{1+}^j = b_1^j \wedge \dots \wedge x_{n+}^j = b_n^j) \end{aligned}$$

Notice that the concentration of the products b_i^j is incremented $\text{dur}(r_j)$ time-units later.

Finally, the whole system looks like this:

$$\text{system} \stackrel{\text{def}}{=} !(\text{state} \parallel \text{choose} \parallel \text{equation}_1 \parallel \dots \parallel \text{equation}_n)$$

The process **state** computes the current concentration of the components according to their concentrations in the previous time-unit and the variables x_{i+}^j and x_{i-}^j explained above. It also sets the initial concentration of the components that are parameters of the simulation. We omit the definition of this process for the sake of presentation.

3.1.1 A simulation tool: BioWayS

We have implemented a simulation tool based on our model, which we have called *BioWayS* and which is freely available on the web. This tool is based on an interpreter of the **ntcc** calculus written in the Oz Languages (<http://www.mozart-oz.org/>).

BioWayS takes as input a text file containing the initial parameters of the simulation as well as the description of the equations. Then, it computes the final stores of each time-unit and outputs the number of equation applied and the concentration of each component. In the following sections we shall show some examples of simulations executed in this tool. More examples and the description of the input and output files can be found at <http://lix.polytechnique.fr/~colarte/bioways/>.

4. IN SILICO EXPERIMENTATION

We used our modelling technique to study some interconnected cellular signaling pathways related to three receptors: Tumor Necrosis Factor Receptor I (TNFR1), Insulin

Receptor (IR), and the Epidermal Growth Factor Receptor (EGFR) [8]. This choice is motivated by the fact that the information regarding these pathways is relatively abundant and can be used to verify the viability of the model comparing them with the output of the simulations. Some aspects of the dynamics of this network have been characterized through *wet-lab* techniques that allow to measure the quantity (concentration or activity) of a pool of cellular proteins over a time period [8].

4.1 The model

The set of signalling pathways composing the network indicated above was considered. Using a system-level approach, our aim was to build a model based on the information available in the literature, adequate for describing the measured data and, eventually, to make predictions regarding the modeled system. Our model network is composed by 37 different kind of molecules (nodes) and 48 equations (interactions). Each interaction is rendered by a “reaction rule” resembling the corresponding chemical equation. The input for a simulation run is composed by the list of rules (i.e. the model of the network) and a set of parameters specifying the number of copies for each involved molecule (i.e. the amount of reacting molecules) and the length of the computation (i.e. the time length of the experiment).

For each rule it is possible to specify a parameter indicating the amount of time needed for the corresponding reaction to occur. This parameter corresponds to the *first passage time* and is evaluated according to [14]. Note that to estimate the first passage time for an enzymatically catalysed reaction, it is not necessary to know the details regarding the reaction’s mechanism and, thus, the kinetic constants related to each elementary step. Instead, the needed data are the enzyme and substrate(s) concentrations, the enzyme-substrate(s) “affinity constant(s)” and the overall (*macroscopic* or *apparent*) rate constant. The first two parameters were taken from the literature or determined heuristically letting their initial amounts ranging on reasonably realistic intervals estimated by comparison with similar cases. The last parameter was estimated using Kinfer (<http://www.cosbi.eu/index.php/research/prototypes/kinfer>) and literature data. The choice of this approach reflects the need of building a “biochemically consistent” model using correctly the available data to avoid common misconceptions, e.g., to use the apparent rate constant instead of the microscopic rate constant. This mistake is often made in works relying on the Gillespie’s Stochastic Simulation Algorithm (SSA) [9] for the (stochastic) simulation of biochemical pathways. Gillespie’s SSA, needs the knowledge of the elementary rate constants, which are often impossible to measure and evaluate. To circumvent this problem, apparent rate constants are typically used. This introduces a heavy approximation with unpredictable effects on the results of simulations.

In our framework, for each step of the computation, *BioWayS* chooses non deterministically one rule and simulates the occurrence of a chemical reaction. When a reaction occurs, reactant molecules are “consumed” and new copies of other molecules are eventually “produced” according to the corresponding rule. After a computation, initial and final states differ for the number of copies of each molecule.

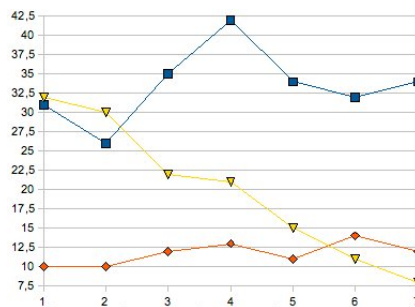


Figure 1: Amount of pAkt present in different experimental conditions: 500I=500μM of insulin (blue); 100T=100μM of TNFα (yellow); 100T 500I=100μM of TNFα and 500I=500μM of insulin (red). This figure should be compared with the homologous in [8]

4.2 Experiments and results

The *in silico* experiments were performed by simulating the model described above through *BioWayS*. In the first experimental session we tuned the model and tested its viability by comparing the simulations outputs with the corresponding wet-lab data reported in [8]. The *in silico* experiments were performed under different initial conditions mimicking the presence of different combinations of insulin, TNFα and EGF in the extracellular environment. For each combination of these three molecules, simulations differing in time length were performed to obtain time courses of the measured values, i.e., the amount of the various kind of molecules at the end of each computation. Some of these results are presented in Figure 1. We compared the “real” and “virtual” time series through the X^2 test ($p > 0.05$). In all the tested cases, the X^2 test showed that our results are consistent with wet-lab data.

For further validating our proposal, we studied particular cases of cross-talk amongst the considered signaling pathways. It is known [3] that the pathways regulated by TNFR1 and IR interact, negatively influencing each other signaling dynamics. The effects of this cross-talk can be observed through proteomic techniques monitoring the amount of phosphorylated Insulin Receptor Substrate 1 (IRS1). IRS1 is specifically phosphorylated on tyrosine and serine residues. Tyrosine phosphorylation of IRS1 is performed by IR when it is bound to insulin and *activates* IR-dependent signaling pathways. Phosphorylation on serine residues decreases the activity of IRS1 and prevents tyrosine phosphorylation. This chemical modification occurs when certain signaling pathways, such as TNFR dependent pathways, are active and leads to the inactivation of IR-dependent signaling pathways. While the relationship between IR and IRS1 is structurally well characterized (IRS1 directly interacts with IR) the same cannot be said for TNFR and IRS1. Actually the model presented in Sec. 4.1 does not reproduce insulin/TNFα interference. This may depend on the fact that in [8] is used a cell culture line (namely human colon adenocarcinoma cells, ATCC) which, being cancerous cells, may not function as the non-cancerous ones such as adipocytes. Our hypothesis was that our model, being shaped on the findings reported in [8], does not mimic “correctly” the insulin-TNFα

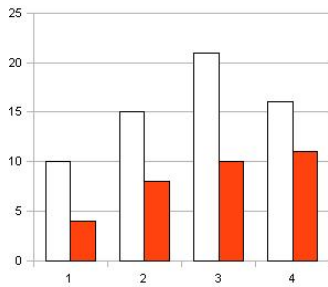


Figure 2: The white columns represents the final amount of IRS1 (arbitrary units) when only insulin is present in the environment. The coloured columns represents the final amount of IRS1 (arbitrary units) when both insulin and $\text{TNF}\alpha$ are initially present.

interactions. Some experimental evidences indicate a possible pathway responsible of this cross-talk [3]. To test our hypothesis, we inserted compositionally this pathway in our model. We then performed a new session of *in silico* experiments by simulating the updated model through *BioWayS* mimicking the presence of insulin, $\text{TNF}\alpha$ or both in the extracellular environment. The results of these experiments (see Figure 2) show that the trend of *in silico* behavior resembles what happens for the real counterpart.

5. RELATED WORK AND CONCLUSIONS

Petri Nets (PN), both in their classical version [10] and in their stochastic ones [15] are one of the first approaches which has been used for modelling living systems. Our formalism should be easier to use as a programming language. Pathway Logic (PL) [16] is a symbolic approach to the modelling and analysis of biological systems which is based on rewriting logic [12]. Our methodology should have the advantage of modelling the time of reactions and treat in a natural way partial information.

The language BIOCHAM was introduced in [6], and was designed for being very close to the classical rules biologists use to describe biochemical reactions. We believe that we can model different sets of reaction rules by exploiting the fact that we do not require precise stoichiometric data. We use constraints for representing partial information.

A programming language similar to ours has been used for representing Biological systems in [2]. [2] considers a language which is suitable for modelling hybrid systems, i.e., system where continuous and discrete time is considered. Some other major differences are as follows. We mainly focus on deriving a simple representation of equations, which can be easily used by non expert users. We model time duration of reactions as well as partial information, like [2], but we present a new methodology which allows to compose several pathways and find results which normally requires the availability of more refined data.

As future work we plan to extend the language and add the possibility to specify the probability to choose a rule, thus introducing the possibility of making stochastic computations as in [11]. We would also like to make use of the declarative nature of *ntcc* as formulae in temporal logic [13] to verify properties of the modelled systems. Another ex-

tension concerns the possibility to model continuous time. This would allow to describe hybrid systems as in [2].

6. REFERENCES

- [1] M.L. Blinov, O. Ruebenacker, and I.I. Moraru. Complexity and modularity of intracellular networks: a systematic approach for modelling and simulation. *IET Syst Biol.*, 2(5):363–368, 2008.
- [2] A. Bockmayr and A. Courtois. Using hybrid concurrent constraint programming to model dynamic biological systems. In *18th International Conference on Logic Programming, ICLP’02*, pages 85–99. Springer LNCS 2401, 2002.
- [3] S. Csehi, S. Mathieu, U. Seifert, A. Lange, M. Zwyer, A. Wernig, and D. Adam. Tumor necrosis factor (tnf) interferes with insulin signaling through the p55 tnfr receptor death domain. *Biochemical and Biophysical Research Communications*, 329:397–405, 2005.
- [4] T.M. Devlin. *Textbook of Biochemistry with Clinical Correlations*. Wiley-Liss, 2002.
- [5] S. Gaudet *et al.* A compendium of signals and responses triggered by prodeath and prosurvival cytokines. *Molecular and Cellular Proteomics*, 4:1569–1590, 2008.
- [6] F. Fages, S. Soliman, and N. Chabrier-Rivier. Modeling and querying interaction networks in the biochemical abstract machine biocham. *Journal of Biological Physics and Chemistry*, 4(2):64–73, 2004.
- [7] T. Finkel and J.S. Gutkind. *Signal Transduction and Human disease*. Wiley-Liss, 2003.
- [8] S. Gaudet, K.A. Janes, J.G. Albeck, E.A. Pace, D.A. Lauffenburger, and P.K. Sorger. A compendium of signals and responses triggered by prodeath and prosurvival cytokines. *Molecular Cellular Proteomics*, 4(10):1569–1590, 2005.
- [9] D.T. Gillespie. Stochastic simulation of chemical kinetics. *Annu. Rev. Phys. Chem.*, 58:35–55, 2007.
- [10] P. Goss and J. Peccoud. Quantitative modeling of stochastic systems in molecular biology by using stochastic petri nets. *Biochemistry*, 95:6750–6754, 1998.
- [11] Vineet Gupta, Radha Jagadeesan, and Vijay A. Saraswat. Probabilistic concurrent constraint programming. In *Proc. of CONCUR 97*, London, UK, 1997. Springer-Verlag.
- [12] N. Martí-Oliet and J. Meseguer. Rewriting Logic: Roadmap and Bibliography. *Theoretical Computer Science*, 285(2):121–154, 2002.
- [13] M. Nielsen, C. Palamidessi, and F. Valencia. Temporal concurrent constraint programming: Denotation, logic and applications. *Nordic Journal of Computing*, 9(1):145–188, 2002.
- [14] J. Ninio. Alternative to the steady-state method: Derivation of reaction rates from first-passage times and pathway probabilities. *Proc. Natl. Acad. Sci. USA*, 84:663–667, 1987.
- [15] V.N. Reddy, M.N. Liebman, and M.L. Mavrovouniotis. Qualitative analysis of biochemical reaction systems. *Computer in Biology and Medicine*, 26(1):9–24, 1996.
- [16] C. Talcott. Pathway logic. *Formal Methods for Computational Systems Biology*, 5016:21–53, 2008.